PRESERVING SPECTRAL CONTRAST IN AMPLITUDE COMPRESSION FOR HEARING AIDS

Juan Carlos Tejero-Calado[1], Janet C. Rutledge[2], and Peggy B. Nelson[3]

[1]University of Malaga, Campus de Teatinos-Complejo Tecnológico, Malaga Spain
[2]Division of Otolaryngology, University of Maryland, Baltimore, MD, USA
[3]Department of Communication Disorders, University of Minnesota, Minneapolis, MN, USA

Abstract-**Amplitude compression processing is used to reduce the amplitude level variations of speech to fit the reduced dynamic ranges of sensorineural impaired listeners. However this processing results in spectral smearing due in part to reduced peak-to-valley ratios. Presented here are two variations of a compression processing algorithm based on a sinusoidal speech model that preserves the important spectral peaks. Both models operate on a time-varying, frequency-dependent basis to adjust to the speech variations and the listener's hearing profile. Preliminary subject tests indicate benefit from preserving spectral contrast. Enhancing spectral contrast is possible with the algorithm presented here.**

I. INTRODUCTION

Sensorineural hearing losses are characterized by a reduced dynamic range of hearing and reduced spectral resolution. Compensating for elevated thresholds involves amplification to raise speech above threshold, while amplitude limiting keeps speech signals from exceeding the impaired listener's threshold of discomfort. Linear techniques apply gain directly to the amplitude of the incoming signal. Peak clipping and compression limiting are used to keep high level signals from exceeding the listener's threshold of discomfort.

Nonlinear techniques, such as amplitude compression, reduce the amplitude level variations of the signal to fit the listener's reduced dynamic range of hearing. Single-channel (wideband) systems process the entire speech signal on the basis of overall level. Multiband syllabic compression systems reduce the variation in speech level in each frequency band according to the subject's reduced dynamic range in that band. However, differing nonlinear processing in adjacent bands can cause audible distortion. The wideband and multiband compression systems mostly use digital or analog filters along with equalization gain. In most cases, compression parameters remain the same over time. Conventional compression processes commonly reduce spectral peak-to-valley ratios in speech, which can hamper speech perception.

Waveform parameterization models, such as sinusoidal modeling, can be used in place of filter-based techniques to achieve linear or compression processing [1,2]. These models allow greater flexibility in the range of compensation processing , and lend themselves to time-varying techniques. Because the sinusoidal model allows manipulation of individual frequency components in each frame, those peaks with the greatest energy in each band can be processed so that their relative shape is maintained. For example, we have described an amplitude compression scheme that preserves the resolution of spectral peaks [1]. This system was "multiband" in the sense that differing amounts of compression were applied to the various frequency regions. However the system avoided the use of physical frequency bands that can lead to distortions from discontinuities at the boundaries between the bands. This scheme is also time varying since the processing is optimized for each frame of speech. The real-time implementation allowed convenient testing on hearing impaired listeners [2].

Presented here is an analysis of the processing algorithm and an alternative implementation that addresses some of the potential shortcomings of the basic model.

II. BASIC PROCESSING ALGORITHM

The sinusoidal model represents speech as the sum of sinusoids with various amplitudes, frequencies and phases. This model has parameters that are independent of voicing state and pitch period. The frequencies of the sinusoids in frame k are chosen to correspond to the $N(k)$ largest peaks in the magnitude of the short-time Fourier transform of the speech signal. The application here, which has been implemented on a TMS320C30 microprocessor, uses 7.5 msec analysis frames and 30 msec Hamming windows, leading to a 4-to-1 time overlap. A 256 point FFT is used to provide sufficient resolution for the speech sampled at 8.013 kHz. Synthesis is done by using an inverse FFT.

The processing algorithm assumes there are up to six important peaks in each frame. The number of peaks selected is determined by the constraint that the peaks must be some minimum frequency spacing from each other. If two spectral peaks are close in frequency, it is assumed that they arise from a single formant. In the example shown in Fig. 1, the top four peaks were chosen, and the processing is optimized for those peaks. The compression ratio is calculated for each principal peak in each frame as the ratio between the impaired and the normal listeners' dynamic range of hearing. These are represented as Δ^* and Δ, respectively. The gain for a given peak is calculated such that the ratio of the peak sensation levels for normal versus

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Preserving Spectral Contrast in Amplitude Compression for Hearing Aids	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) University of Malaga, Campus de Teatinos-Complejo Tecnológico, Malaga Spain	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the 23rd Annual International Conference of the IEEE ENgineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

impaired listeners is equal to the ratio of their respective dynamic ranges. In other words,

$$c = \frac{\Delta^*}{\Delta} = \frac{\delta^*}{\delta} \qquad (1)$$

where Δ is the normal dynamic range and Δ^* is impaired dynamic range at the frequency of the peak, δ is the sensation level for the sinusoid for normal listeners, and δ^* is the sensation level for the processed sinusoid for impaired listeners. The amplitude of the processed peak is

$$A^* = cA + T_{im} \qquad (2)$$

where, T_{im} is the impaired threshold of hearing, and A is the amplitude of the original sinusoid peak. The resulting gain is the difference between A^* and A.

Fig. 1 shows the spectrum of a speech segment. The peaks of the FFT that are used as the underlying sinusoids are indicated by a $*$. The squares denote the sinusoids that are selected as principal peaks. The sloping line represents the threshold of hearing T_{im} for the impaired listener used in this example.

In order to preserve the relative peak-to-valley ratio of the principal peaks, the gains for those peaks are used to determine the amount of gain to apply to the sinusoids between the principal peaks. To prevent discontinuities, the gain applied to each sinusoid transitions linearly from the gain at one principal peak to the gain at the next principal peak. The resulting speech has the reduced dynamic range characteristic of compression processing even though most of the peaks actually undergo linear gain processing. Henceforth this processing scheme will be referred to as COL to represent its compression and linear gain aspects.

The processed speech is synthesized by performing an inverse FFT on the modified sinusoid peaks. The processing operates on the magnitudes of the complex spectral amplitudes only; the phase is not changed. The quality of the algorithm relies on a naturally changing phase, and not on the exact values of the phase components. The 4-to-1 overlap-add process used in the synthesis smoothes discontinuities at the frame boundaries. Although the peak-to-valley ratio is not perfectly maintained, it is evident from Fig. 1 that the important peaks are clearly distinguishable in the COL processed speech and do not suffer from the smearing that accompanies conventional compression processing.

To reduce ambient background noise in silent regions, if the maximum peak in a band is below 30 dB, it is only given the gain that would be applied to a 30-dB peak. Some gain is needed in these areas to avoid the perception of discontinuities in the signal. It is presumed, however , that the low-level peaks are not informative parts of the speech signal. This parameter is adjustable and can be optimized based on the results of clinical tests.
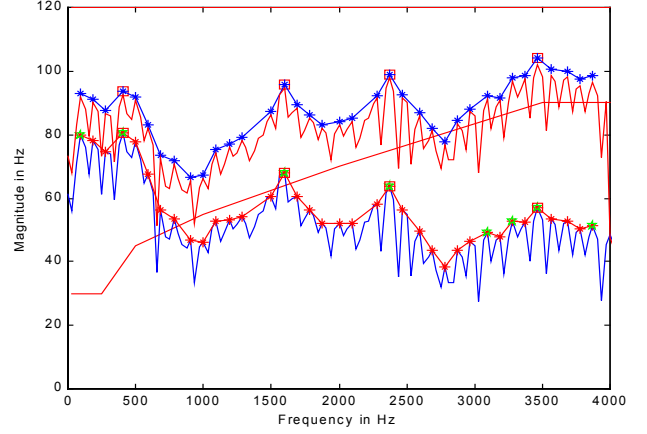

Fig. 1. Speech spectrum indicating top four spectrally important principal peaks before and after COL processing. The impaired threshold of hearing T_{im} is also shown. The peaks of the FFT used as the model sinusoids are indicated by a $*$. The squares denote the sinusoids that are selected as principal peaks

III. RESULTS AND ANALYSIS

The theoretical benefits of preserving and enhancing speech spectral contrast for listeners with impaired hearing have been outlined previously [3]. The results obtained with the COL processing scheme from hard of hearing listeners show that preserving spectral peak-to-valley ratios can improve word and sentence understanding in quiet and in noise [4]. The COL algorithm was compared to a multiband compression (MBC) system with five bands and a 40 dB threshold similar to those compression systems implemented in current hearing aids. Stimuli processed by both algorithms had the same long-term spectrum and overall amplitude, but contained different spectral peak-to-valley ratios. Stimuli were presented at a range of intensities in both quiet and in noise.

Fig. 2 shows the spectra for COL and MBC processing along with the original spectrum for a segment of speech. Both processing methods clearly raise the speech to an audible range. Although the MBC signal (top spectrum) for this frame appears stronger, all processed signals were matched for rms prior to presentation. Note the greater peak-to-valley contrast for the COL signal (middle spectrum). Note also some discontinuities in the MBC signal that occur at the boundary between frequency bands.

Testing was conducted on COL and MBC processing using four adult listeners with moderate hearing loss (flat and sloping). Stimuli consisted of consonant-vowel-consonant (CVC) nonsense syllables under four listening conditions: in quiet and in noise, each at both soft and comfortable levels. Here "soft" is defined as 10 dB below the most comfortable level (MCL). Results are shown in Table 1 for the both soft and MCL listening levels in both quiet and in noise.
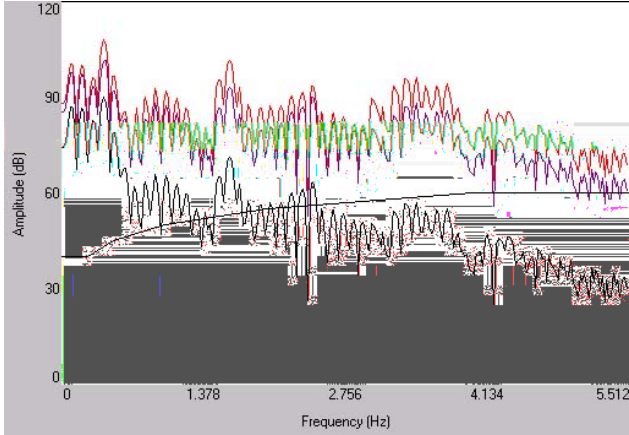
Fig. 2. Speech spectra for COL and MBC processing schemes. The top spectrum is MBC, COL is just below that, and the original spectrum is on the bottom. The impaired threshold of hearing T_{im} is also shown.

At soft levels, listeners tended to perform better using COL, especially in quiet. At comfortable levels, listeners also tended to perform better using COL in noise. However at comfortable levels with no noise, both processing methods performed equally well. Overall, all of the hearing impaired listeners, either flat or sloping losses, demonstrated benefit from COL processing.

The main observable differences in the outputs of the two processing schemes are in the spectral resolution, and the tendency for MBC to have higher amplitude levels in high frequency regions for sloping loss subjects. The reasons for the differences in spectral resolution have already been noted. The COL processing tends to select primary peaks in the lower frequency regions so the gains for high frequency regions are often driven by what was determined in a region of lesser loss. When the COL is forced to choose a primary peak above 4 kHz, the output spectrum is closer to that of MBC in that frequency region. However in many speech segments there is not much informative speech signal information above 4 kHz so it is not clear whether this slight change in amplitude level at high frequencies is having a pronounced effect. In preliminary listening tests using only COL, there did not appear to be any difference in performance when a primary peak was forced to be above 4 kHz. Recent clinical data suggest that providing amplification in the region of 4 kHz is often undesirable for listeners with severe hearing loss [5]. However more tests need to be conducted.

The question of whether signal processing schemes that preserve or enhance spectral contrast can compensate for the reduced frequency resolution was further examined in [4]. Normal hearing listeners with simulated hearing losses were used in tests similar to the ones conducted with hearing impaired listeners to determine whether the benefit observed previously was primarily from increased audibility or from improved spectral resolution. Results were also compared with those obtained using original unprocessed speech.

In summary, speech processed through COL was as intelligible as the original for both simulated mild and sloping hearing losses. In fact final consonants may have been more intelligible with COL than with the original. Speech processed through MBC appropriate for mild losses was as intelligible as the original. However speech processed through MBC for sloping losses was less intelligible than the original or COL. MBC for this hearing loss introduces some artifacts and spectral smearing. Overall, the improvement for normal hearing listeners was not as great as that noted by listeners with hearing loss in the previous study. Therefore it can be concluded that the increased audibility provided by COL is not the main benefit. Rather the results suggest that the improved spectral contrast with COL provides some compensation for the reduced spectral resolution in hearing impaired listeners.

Table 1. Mean *(and standard deviation)* number of phonemes correct for 4 hard-of-hearing listeners identifying consonants and vowels in 20 syllables processed by COL or by MBC. Syllables were presented at the listeners' most comfortable levels (MCL) or at soft levels 10 dB below their MCL, in broadband noise or with no noise added. Asterisks (*) indicate a significant difference between scores obtained using COL and MBC processing.

	COL		MBC	
	Consonants	Vowels	Consonants	Vowels
Soft level-noise	4.3* *(1.3)*	8.3 *(2.6)*	2.8 *(1.8)*	6.4 *(1.5)*
Soft level no noise	11.4* *(3.9)*	15.3* *(3.6)*	8.5 *(4.0)*	11.8 *(3.4)*
MCL-noise	6.8* *(2.6)*	13.8 *(2.2)*	5.0 *(1.7)*	9.0 *(4.1)*
MCL-no noise	11.6 *(.4)*	13.8 *(1.0)*	9.8 *(4.7)*	13.0 *(3.8)*

IV. LPC-BASED IMPLEMENTATION OF ALGORITHM

The basic COL algorithm selects the primary peaks through an iterative process that eliminates certain peaks from further consideration. The selected peaks are determined in part by the maximum number of peaks allowed and the minimum spacing that must be between them. The values used here were determined mostly through experimental means.

To test further the importance of spectral resolution in compression processing, it is desirable to enhance the peak-to-valley ratio present in the original speech, rather than simply preserving it. This requires more precision in selecting the principal peaks. In fact, theory suggests that we choose peaks at the formant frequencies rather than just at some relative maxima of the spectrum. Using the LPC spectrum as a guide is one convenient way of achieving this goal. It also ensures that the principal peaks will be appropriately distributed throughout the spectrum, and not clustered in lower frequency, higher amplitude areas.

The COL-LPC processing is similar in principle to the COL processing. However rather than operating on the sinusoidal model peaks, the algorithm finds the desired gain for each principal peak of the LPC spectrum, and then interpolates the LPC valley regions between peaks in the

same manner as with COL. The gain for each sinusoidal peak is then chosen to match the processed LPC spectrum.

Fig. 3 shows the selected peaks and processed speech for both the COL and COL-LPC processing. In this example, the COL algorithm chose a principal peak that is actually in a valley region since it is a local maximum and is the required distance from another chosen peak. It also tends to favor lower frequency regions. In contrast, the COL-LPC will only select principal peaks that are actual peaks in the spectrum (formants). There are only slight differences in the spectra; however, the higher frequency portion of the COL-LPC spectrum is slightly higher in amplitude. These differences are very difficult to detect in listening tests.

V. Discussion and Conclusions

The characteristics of sensorineural hearing loss, particularly in the case of sloping losses, suggest that amplitude compression processing could benefit many hearing impaired listeners. To date, most of the test results with conventional multiband processing have been disappointing. This may be due to the implementation of multiband processing, and not its inherent properties. Presented here are two variations of a processing algorithm that exploit the strengths of the sinusoidal model to produce high quality amplitude compressed speech that is within the residual dynamic range of the impaired listener and has strong resolution of the important spectral peaks. Both operate on a time-varying basis to adjust to the characteristics of the speech in each frame, and on a frequency-dependent basis to accommodate the shape of the hearing loss.

Another potential benefit of both COL and COL-LPC is that they allow fast compression without the discontinuities that occur in filter-based systems. The basilar membrane accomplishes its compression instantaneously which suggests that an effective processing system should come close to instantaneous compression. Because of the distortions that accompany multiband filter-based systems, listeners tend to prefer longer time constants that cause fewer audible changes in the signal.

Testing has not been conducted on COL-LPC, however, we expect it to perform similarly to the basic COL algorithm when no additional spectral enhancement is performed. The flexibility of COL-LPC will allow the unique opportunity to conduct further research on the benefits of spectral enhancement in combination with compression processing. Future plans call for extensive testing under a variety of listening conditions using the real-time system.
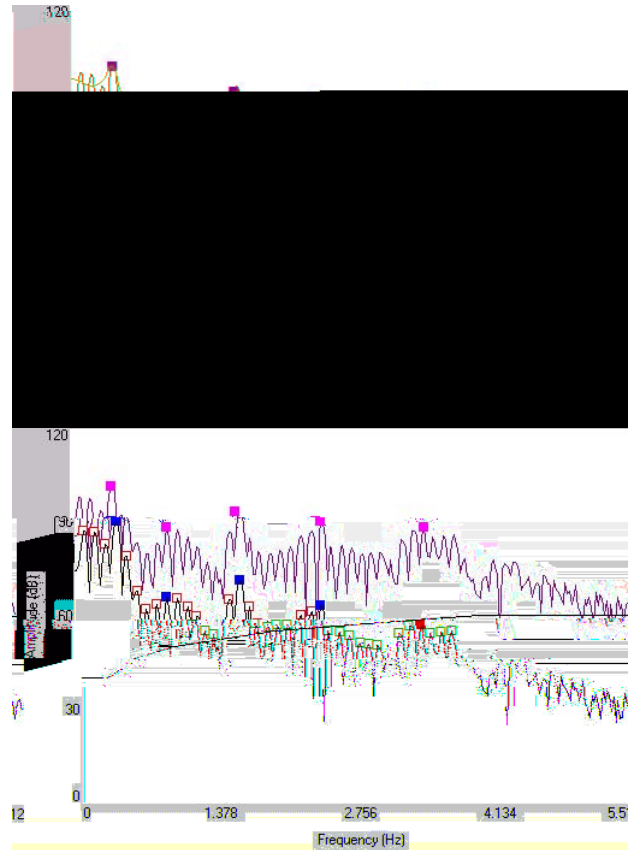
Acknowledgments

Fig. 3. Principal peaks and spectra for COL-LPC (top panel) and COL (bottom panel). COL-LPC spectrum is overlaid with the LPC spectrum

References

[1] J.C. Tejero-Calado, J.C. Rutledge and P.B. Nelson, "Combination compression and linear gain processing for digital hearing aids," in *Proceedings of 20th Annual International Conference of the IEEE Engineering in Medicine & Biology Society,* October 1998.

[2] J.C. Tejero-Calado, J.C. Rutledge and P.B. Nelson, "Real-time combined compression and linear gain processing for digital hearing aids," in *Proceedings of the Joint Meeting of the 137th Meeting of the Acoustical Society of America and 2nd Convention of the European Acoustics Association*, March 1999.

[3] R. Miller, B. Calhoun and E. Young, "Contrast enhancement improves the representation of /e/-like vowels in the hearing-impaired auditory nerve," *Journal of the Acoustical Society of America,* vol. 106(5), 1999, pp. 2693-2708.

[4] P.B. Nelson, J.C. Tejero-Calado, J.C. Rutledge, "Benefits of preserving speech spectral contrast for listeners with hearing loss," *Association for Research in Otolaryngology 2001 MidWinter Meeting,* February 2001.

[5] C.W. Turner and K.J. Cummings, "Speech audibility for listeners with high-frequency hearing loss," *American Journal of Audiology,* vol. 8(1), 1999, pp. 47-56.